732A38 Computational statistics
Division of Statistics and Machine Learning
Department of Computer and Information Science

Computer lab 6

Instructions

- Create a report to the lab solutions in PDF.
- Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
- **Include all your codes as an appendix into your report.**
- A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes.
- The lab report should be submitted via LISAM before the deadline.

Assignment 1: Genetic algorithm

In this assignment, you will try to perform one-dimensional maximization with the help of a genetic algorithm.

1. Define the function $f(x) = \frac{x^2}{e^x} - 2e^{\frac{-9\sin(x)}{x^2+x+1}}$.
2. Define the function "crossover" that for two scalars *x* and *y* returns their "kid" as *(x+y)/2*
3. Define the function "mutate" that for scalar *x* returns the result of the integer division x^2 *mod* 30. (Operation "mod" is denoted in R as "%%")
4. Write a function that depends on the parameters *maxiter and mutprob* and:
    a. Plots function *f* in the range from 0 to 30. Do you see any maximum value?
    b. Defines an initial population for the genetic algorithm as X=(0,5,10,15,…30)
    c. Computes vector "Values" that contains the function values for each population point
    d. Performs *maxiter* iterations where at each iteration
        i. Two indexes are randomly sampled from the current population, they are further used as parents (use *sample().*)
        ii. One index with the smallest objective function is selected from the current population, the point is referred to as victim(use *order()*)
        iii. Parents are used to produce a new kid by crossover. Mutate this kid with probability *mutprob*. (use *crossover(), mutate()*)
        iv. The victim is replaced by the kid in the population and the list *"Values"* is updated
        v. The current maximal value of the objective function is saved

   e. Final observations are added to the current plot and marked by some other color.
5. Run your code with different combinations of *maxiter=10, 100* and *mutprob=0.1, 0.5,0.9.* Observe the initial population and final population. Conclusions?

Assignment 2: EM algorithm

Data file ***physical.csv*** describes a behavior of two related physical processes $Y = Y(X)$ and $Z = Z(X)$

1. Make a time series plot describing dependence of Z and Y versus X. Does it seem that two processes are related to each other? What can you say about the variation of the response values with respect to X?
2. Note that there are some missing values of Z in the data which implies problems in estimating models by maximum likelihood. Use the following model

$$Y_i \sim Exp\left(\frac{X_i}{\lambda}\right), Z_i = Exp\left(\frac{X_i}{2\lambda}\right)$$

   where λ is some unknown parameter to derive an EM algorithm that estimates λ.

3. Implement this algorithm in R and use $\lambda_0 = 100$ and convergence criterion "stop if the change in λ is less than 0.001". What is the optimal λ and how many iterations were required to compute it?
4. Plot EY and EZ versus X in the same plot as Y and Z versus X and comment whether the computed λ seems to be reasonable.

Submission procedure

Assume that X is the current lab number.

If you are neither speaker nor opponent for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members submits the group report using *Lab X group report* in the *Submissions* folder before the deadline

If you are a speaker for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members does the following before the deadline:

- submits the group report using *Lab X group report* in the *Submissions* folder before the deadline
- Goes to Study room *Speakers X→Documents* and opens file *Password X.txt*. Then the student should put your group report into ZIP file *Lab X.zip* and protect it with a password you found in Password X.txt
- Uploads the file to *Collaborative workspace* folder

If you are opponent for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members submits the group report using *Lab X group report* in the *Submissions* folder before the deadline
- After the deadline for the lab has passed, go to Collaborative workspace folder and download *Lab X.zip*. Open the PDF in this ZIP file by using the password available in *Course Documents→Password X.txt,* read it carefully and **prepare at least two questions/comments/improvement suggestions** in order to put them at the seminar (i.e. at least two questions per opponent)